



Electronic data publication in geochemistry: A plea for “full disclosure”

Hubert Staudigel

Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego – 0225, La Jolla, California 92093-0225, USA (hstaudigel@ucsd.edu)

Francis Albarede

Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon, cedex 7 France (albarede@ens-lyon.fr)

Don L. Anderson

Seismological Laboratory, California Institute of Technology, MC 252-21, Pasadena, California 91125, USA (dla@gps.caltech.edu)

Louis Derry

Department of Geological Sciences, Snee Hall, Cornell University, Ithaca, New York 14853, USA (derry@geology.geo.cornell.edu)

Bill McDonough

Department of Geology, University of Maryland, College Park, Maryland 20742, USA (mcdonough@geol.umd.edu)

Henry F. Shaw

Division of Chemical Sciences, Geosciences, and Biosciences, SC-14, Office of Basic Energy Sciences, U.S. Department of Energy, 19901 Germantown Road, Germantown, Maryland 20874-1290, USA (Henry.Shaw@hq.doe.gov)

William White

Department of Geological Sciences, Snee Hall, Cornell University, Ithaca, New York 14853, USA (white@geology.cornell.edu)

Alan Zindler

National High Magnetic Field Laboratory and Department of Geological Sciences, Florida State University, 1800 East Paul Dirac Drive, Tallahassee, Florida 32306-4005, USA (zindler@magnet.fsu.edu)

Keywords: Databases; data formats; Geochemical Earth Reference Model.

Index terms: Geochemistry: Instruments and techniques; Hydrology: Instruments and techniques; Mineralogy and petrology: Instruments and techniques.

Received August 23, 2001; **Revised** September 11, 2001; **Accepted** September 12, 2001;

Published October 11, 2001.

GERM Steering Committee, 2001. Electronic data publication in geochemistry: A plea for “full disclosure,” *Geochem. Geophys. Geosyst.*, vol. 2, Paper number 2001GC000234 [1026 words]. Published October 11, 2001.

[1] Computer technology and automated analytical instruments have resulted in an explosion in the quantity of geochemical data produced over the past several decades. While this is an extremely positive development for science, it has had the unfortunate effects of diminishing the value of individual analyses and has created a reluctance on the part of authors, editors, and publishers to publish data. Data, it seems, like much of the rest of what our society produces, have become disposable. After a short appearance in glossy illustrations, the fate of analytical results is to be buried, if not in landfills, then at least on shelves in investigators' offices.

[2] One man's castaway can be another man's treasure. Data are useful well beyond the initial publication in a scientific paper. One might argue that interpretations of data may change and the data are the only permanent aspect of a paper. There are a number of important scientific papers based entirely, or almost so, on previously published data. If, however, the data are not published, such contributions are not possible. There is a good chance that geochemical samples have to be reanalyzed simply because data are not preserved for future use, beginning a new cycle of data generation and burial. Data are a most important heritage for geochemistry that has to be preserved for the future. Therefore our plea is for full publication of all data contributing to papers in geochemistry.

[3] This is already the policy of the American Geophysical Union (AGU). The policy adopted by AGU's publications committee in 1993 states "Data sets discussed in data papers published in AGU books and journals must be publicly available and accessible to the scientific community indefinitely," and "data sets that are available only from the author, through miscellaneous public network services, or academic, government or commercial insti-

tutions not chartered specifically for archiving data, may not be cited in AGU publications" (available at http://www.agu.org/pubs/data_policy.html). Our plea, then, is that authors and editors respect this policy and that other publishers adopt a similar one.

[4] Computer technology, which has largely led to the data "devaluation" mentioned above, also provides the potential solution through information technology (IT). Like most other earth science disciplines, geochemistry stands to reap substantial benefits from embracing IT. These benefits include wider dissemination of geochemical data, increased ease of use of data by different earth science subdisciplines, and more efficient storage and retrieval of archived data. In the "Gutenberg" age, data had to be transmitted on the printed page. The cost of doing so was the main reason for the reluctance to publish large volumes of data. In the current "electronic" age communicating data this way is no longer necessary. Furthermore, it is not desirable. For most of us, data are most useful in our computers, where we can analyze them, compare results, create visualizations, etc.

[5] Our plea is, thus, a plea for full electronic publication of data. Most publishers now make electronic supplements available through the World Wide Web. This is certainly a positive development, and it is hoped that publishers commit to permanently archive these data. Authors, however, need to make use of these supplements, and editors need to insure that they do. Publishers need to make these web sites easy to use and readily accessible.

[6] Electronic publication of analytical results, however, does not fully solve the problem. There are currently no standards for the submission of data and metadata, which substantially limits the utility of many data sets. Data formats do not have to be restrictive, but they have to be clearly described, so differing units

or normalizations can be transformed into a common format. Metadata, data about data, should include a full characterization of the samples analyzed and the precise location where the sample was taken and there should be full disclosure of analytical techniques and sample processing. Such data can be coded as text. Definition of data and metadata formats in electronic data supplements has to define new standards of scholarly data publications, but they do not have to be rigid. They need to

allow for expansion as science proceeds or omission when particular metadata types are no longer useful.

[7] The Geochemical Earth Reference Model (GERM) Initiative has made some proposals toward definition of data and metadata formats that is open to discussion and community input (<http://earthref.org/metadata/GERM/>). We invite suggestions or alternative proposals from the geochemical community.